University of Southern Denmark
IMADA
Arthur Zimek
Jonatan Møller Gøttcke, Jonas Herskind Sejr

DM566/DM868/DM870: Data Mining and Machine Learning

Spring term 2019

Exercise 3: Apriori, Confidence, Closed Frequent Itemsets

Exercise 3-1 Apriori candidate generation

Given the frequent 3-itemsets:

$$\{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,3,4\},\{1,3,5\},\{2,3,4\},\{2,3,5\},\{3,4,5\}$$

List all candidate 4-itemsets following the Apriori joining and pruning procedure.

Exercise 3-2 The monotonicity of confidence

Theorem 2.1 in the Lecture states:

Given:

- itemset X
- $Y \subset X, Y \neq \emptyset$

If $\text{conf}(Y \Rightarrow (X \setminus Y)) < c$, then $\forall Y' \subset Y$:

$$\text{conf}(Y' \Rightarrow (X \setminus Y')) < c.$$

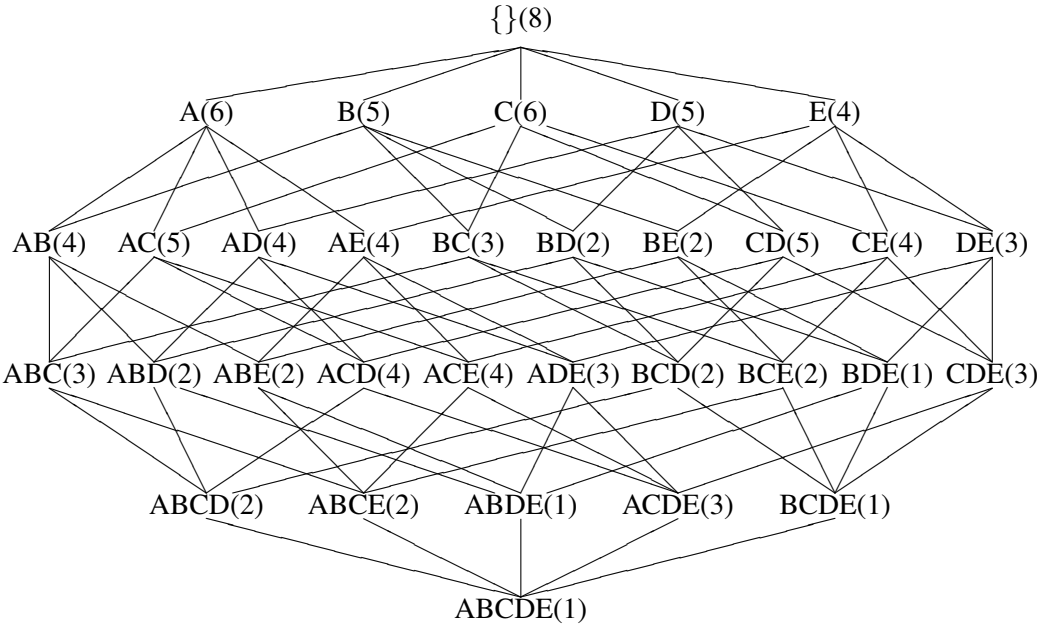
(a) Prove the theorem.

(b) Sketch an algorithm (pseudo code) that generates all association rules with support σ or above and a minimum confidence of c, provided the set F of all frequent itemsets (w.r.t. σ) with their support, efficiently using the pruning power of the given theorem.

Exercise 3-3 Support based on closed frequent itemsets

(a) The database from the lecture grew by one transaction. We computed the corresponding support of all itemsets in the lattice:

TID	A	B	C	D	E
1	0	1	0	0	0
2	1	0	1	1	1
3	1	1	1	0	1
4	0	0	1	1	0
5	1	1	1	1	1
6	1	0	1	1	1
7	1	1	0	0	0
8	1	1	1	1	0



Identify the closed frequent itemsets for the support thresholds $\sigma = 4$ and $\sigma = 2$, respectively.

What do you observe?

(b) Sketch an algorithm (pseudo code) to find the support for all frequent itemsets, using only the set of closed frequent itemsets as information.